Investigation of the mechanism of the SpnF-catalyzed [4+2]-cycloaddition reaction in the biosynthesis of spinosyn A

Byung-sun Jeon*, Mark W. Ruszczycyba,1, William K. Russellb, Geng-Min Linb, Namho Kimb, Sei-hyun Choa, Shao-An Wanga, Yung-nan Liub, John W. Patrickc, David H. Russelld, and Hung-wen Liua,b,1

*Department of Chemistry, University of Texas at Austin, Austin, TX 78712; bDivision of Chemical Biology and Medicinal Chemistry, College of Pharmacy, University of Texas at Austin, Austin, TX 78712; and cDepartment of Chemistry, Texas A&M University, College Station, TX 77843

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved August 7, 2017 (received for review June 12, 2017)

The Diels–Alder reaction is one of the most common methods to chemically synthesize a six-membered carbocycle. While it has long been speculated that the cyclohexene moiety found in many secondary metabolites is also introduced via similar chemistry, the enzyme SpnF involved in the biosynthesis of the insecticide spinosyn A in Saccharopolyspora spinosa is the first enzyme for which catalysis of an intramolecular [4 + 2]-cycloaddition has been experimentally verified as its only known function. Since its discovery, a number of additional standalone [4 + 2]-cycloadditions have been reported as potential Diels–Alderases; however, whether their catalytic cycles involve a concerted or stepwise cyclization mechanism has not been addressed experimentally. Here, we report direct experimental interrogation of the reaction coordinate for the [4 + 2]-carbocyclase SpnF via the measurement of α-secondary deuterium kinetic isotope effects (KIEs) and measured the corresponding secondary deuterium KIEs on both the nonenzymatic and enzyme-catalyzed cyclization of the SpnF substrate. The measured KIEs for the nonenzymatic reaction are consistent with previous computational results implicating an intermediary state between formation of the first and second carbon–carbon bonds. The KIEs measured for the enzymatic reaction suggest a similar mechanism of cyclization within the enzyme active site; however, there is evidence that conformational restriction of the substrate may play a role in catalysis.

Diel-Alderase | isotope effect | enzyme | spinosyn | Bayesian analysis

The enzyme SpnF from Saccharopolyspora spinosa catalyzes an intramolecular [4 + 2]-cycloaddition (4 → 5) during biosynthesis of the insecticide spinosyn A (1) (Fig. 1) (1, 2). At the time of its discovery, the search for a naturally selected Diels–Alderase had already led to the characterization of several enzymes proposed to catalyze [4 + 2]-cycloadditions (3). In all cases, however, these potential Diels–Alderases were found to be multifunctional, often catalyzing a priming reaction that activates the substrate toward cyclization. Thus, the detailed mechanisms of cyclization and the catalytic roles played by the enzymes were unclear and difficult to address. In contrast, the priming reaction (3 → 4) during spinosyn A biosynthesis is catalyzed by the separate enzyme SpnM (1). The resulting 1, 4-dehydration product 4 can then undergo stereospecific cyclization in aqueous solution with a half-life of ~25 min at 300 K; however, in the presence of SpnF, a rate enhancement (i.e., kcat/km Rm) of ~500-fold is observed, making SpnF the first reported example of a standalone [4 + 2]-carbocyclase (1). This separation of activities has made SpnF an ideal system for mechanistic study.

Since the discovery of SpnF, a number of additional enzymes, such as VtsJ (4), PyrH (5), PyrE3 (5), TecM (6), TbtD (7), and AbYU (8), have been characterized as uniquely catalyzing [4 + 2]-cycloadditions; however, investigation of these enzymes has been limited primarily to crystallographic and computational studies in large part because of the structural complexity of their substrates and the limited options available to study the mechanisms of biological cycloaddition reactions (3). For the purposes of the following discussion, we regard an intermediate as a local minimum in the free energy along the reaction coordinate and a concerted cyclization as one lacking any such intermediate. Consequently, there has yet to be a direct experimental interrogation of the reaction coordinates that defines whether the catalytic cycles of these [4 + 2]-carbocyclases proceed via concerted or stepwise mechanisms.

Development of a convergent synthetic strategy (2) has made it possible to prepare isotopologs of the SpnF substrate, allowing study of the SpnF reaction via the measurement of secondary deuterium kinetic isotope effects (KIEs). Previous chemical model studies have indicated that concerted formation of both C–C bonds in a [4 + 2]-cycloaddition reaction (e.g., the C4–C12 and C7–C11 bonds in 1) should result in inverse deuterium KIEs of about 5% at all four carbons undergoing sp2 to sp3 rehybridization (9, 10). In contrast, if cyclization is stepwise, with the first step most rate-limiting, then unit to normal isotope effects would be expected for carbons involved in the second bond formation (9, 11). Since a stepwise intramolecular cyclization for the SpnF-catalyzed cycloaddition (Fig. 2) is expected to proceed with C7–C11 bond formation first, the C4D and C12D secondary deuterium KIEs are of particular interest in determining the sequence of bond formation. With this goal in mind, we prepared four site-specifically deuterated isotopologs of 4 (C4D, C7D, C11D, and C12D) and measured the corresponding secondary deuterium KIEs on both the nonenzymatic and enzyme-catalyzed reactions. The C2D and C14D isotopologs were also

Significance

Computational and structural studies of putative Diels–Alderases and the reactions that they catalyze have provided testable models of the reaction coordinates and mechanisms of catalysis; however, there has yet to be a direct experimental interrogation of these hypotheses. Herein, the α-secondary deuterium kinetic isotope effects at all sites of rehybridization on the nonenzymatic and SpnF-catalyzed [4+2]-cycloaddition during biosynthesis of spinosyn A are reported. Results implicate an intermediary state between formation of the first and second carbon–carbon bonds and provide evidence that conformational restriction of the substrate may play a role in catalysis.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence may be addressed. Email: mwr8@utexas.edu or h.w.liu@mail.utexas.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1710496114/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1710496114

PNAS Early Edition | 1 of 6
prepared to assess the possible involvement of these two carbon centers in the cyclization reaction as predicted by a recent computational investigation (12).

**Results and Discussion**

The site-specific KIEs were measured using the method of internal competition (14) and whole-molecule MS to monitor changes in the isotopolog enrichment of residual substrate (15–17). As discussed in *SI Appendix, SI Text*, this approach requires using isotopically enriched substrates to enhance observable changes in isotopolog populations and minimize any measurement bias caused by natural abundance labeling. Natural abundance and site-specifically deuterated isotopologs of the macrolactone 2 were prepared by connecting the three fragments A, B, and C via two C–C bond coupling reactions as shown in Fig. 3. A Julia–Kocienski olefination between fragments A and B yielded the Δ12-(E) olefin, and a palladium-catalyzed Stille coupling with fragment C led to the C-5/C-6 linkage. Subsequent cyclization of the resulting linear polyketide by Yamaguchi macrolactonization completed the assembly of 2. The corresponding SpnM substrate (3) was then prepared and isolated from reactions with the biosynthetic enzyme SpnJ (18). This modular synthetic approach facilitated preparation of the six isotopologs of 3 by incorporating the appropriately deuterated fragments (A–C) as detailed in *SI Appendix, SI Text*.

Addition of SpnM to a mixture of natural abundance 3 and a site-specifically deuterated isotopolog in H2O/DMSO at 30 °C generated the isotopically enriched 4 in situ. After complete consumption of 3, residual 4 was isolated by HPLC from six reaction aliquots spanning ~10–90% conversion of 4 to 5 in the presence of sufficient SpnF to ensure that reaction flux is predominantly through the enzyme-catalyzed reaction. Parallel experiments were also performed in the absence of SpnF to study the nonenzyme-catalyzed cyclization. The observed isotopolog ratio of each isolated sample of residual SpnF substrate (4) was then determined by electrospray ionization time-of-flight MS by taking the ratio of the \( M+1 \) and \( M \) peak intensities for the sodiated ion of 4 and correcting for natural abundance labeling (19). Experimental and analytical details are provided in *Materials and Methods*. At least three experimental trials, each consisting of approximately six isolated samples of residual 4, were performed for all site-specific KIEs measured.

Since the expected secondary KIEs are all near unity, the ratio of isotopolog concentrations in the residual starting material (heavy over light) can be expressed as a function \( R \) of the fraction of reaction \( f \) parameterized in terms of the isotope effect \( Dk \) and initial enrichment ratio \( \rho \) according to (20).

\[
R(f; Dk, \rho) = \rho (1 - f)^{1/Dk - 1}, \quad 0 \leq f < 1.
\]

We used a hierarchical Bayesian regression analysis using Markov chain Monte Carlo (MCMC) (21, 22) rather than least squares fitting of Eq. 1 to model uncertainty in the measured KIEs caused by both within- and between-experiment variation while at the same time, estimating posterior probability distributions for the KIEs. Details regarding this analysis, including construction of priors, are described in *Materials and Methods*. Fig. 4 lists posterior point estimates of the site-specific KIEs as well as examples of enrichment plots for the C4D and C12D labeling patterns. The full dataset and simulated posterior marginal probability distributions of the KIEs are provided in *SI Appendix, Figs. S3 and S4 and Tables S2 and S3*.

Given a quasisteady state, the KIE on the cyclization of 4 to 5 in \( n \) elementary reactions (steps) can be expressed as the weighted average

\[
D_k = \sum_{i=1}^{n} S_i D_{k_i},
\]

where \( D_{k_i} \) is the effective KIE associated with progression from the reference state (i.e., 4 free in solution) to the \( i \)th transition state, and \( S_i \) is Ray’s sensitivity index for the \( i \)th step (23–25). Each sensitivity index can be written as a reciprocal sum of forward and reverse commitments and reflects the degree to which a step may be considered “rate-limiting” (24, 25). Applied to the nonenzymatic reaction, the roughly 5% inverse values for

---

**Fig. 1.** Biosynthetic pathway of spinosyn A in *S. spinosa* (1, 2). SpnA–E together constitute a polyketide synthase (PKS) system.

**Fig. 2.** Mechanistic proposal for the [4+2]-cycloaddition of the SpnF substrate 4. Lewis structures represent putative local minima (intermediates) and maxima (transition states) along the generalized free energy curves of the bifurcating reaction coordinate.
than 0.90 is no greater than 10% (SI Appendix, Fig. S3). Therefore, the reduced magnitude of \( C_{TD} k_{non} \) and \( C_{11D} k_{non} \) vs. the computed values suggests that the actual transition state (represented by \( 6 \)) is likely earlier than that predicted by the computational gas-phase model. The KIEs are not consistent, however, with a pericyclic \([4 + 2]\)-transition state for a concerted concerted cyclization, which would predict inverse effects at all four sites of \( sp^2 \rightarrow sp^3 \) rehybridization (9, 10, 29). In so far as the 10π-aromatic state can be considered an intermediate following these observations suggest that nonenzymatic cyclization of 4 may indeed be stepwise, albeit not in the usual sense of a biradical or dipolar intermediate. It should be emphasized, however, that the KIE measurements do not address existence of the \([6 + 4]\)-cycladduct (9) along the reaction coordinate, because they only provide information regarding the first, more rate-limiting elementary reaction.

The enzymatic KIEs can also be interpreted in terms of Eq. 2, where \( k_{enz} \) corresponds to the isotope effect on \( V/K \) for SpnF (14, 25). If substrate binding were the most rate-limiting (i.e., \( S_1 \approx 1 \)), then roughly unit isotope effects would have been expected at all six positions. If product dissociation were the most rate-limiting (i.e., \( S_n \approx 1 \)), then the observed KIE would reflect the equilibrium isotope effect on \( sp^2 \rightarrow sp^3 \) rehybridization (23, 25), and four of six KIEs would be at least 5% inverse regardless of whether the product released is a \([4 + 2]\)- or \([6 + 4]\)-cycladduct (9, 29). That these outcomes were not observed suggests that the intermediary steps of the catalytic cycle dominate the observed enzymatic isotope effects.

Unlike the nonenzymatic cyclization, computational study of the SpnF-catalyzed reaction has been hampered, because a structure of the SpnF–substrate complex is not yet available. Nevertheless, the enzymatic results largely mirror those for the nonenzymatic reaction with the exception of \( C_{14D} k_{non} \), which is notably inverse, whereas the value of \( C_{12D} k_{non} \) is comparable with that of the nonenzymatic case as shown in Fig. 4. While the posterior marginal probability that \( C_{14D} k_{non} < 0.98 \) is greater than 90%, the ~5% normal value for \( C_{12D} k_{non} \) argues against significant C4–C12 bond formation in the transition state of the most rate-limiting step. Structural investigations of the \([4 + 2]\)-carbocycles SnF, AbyU, and PyrE3 all implicate the formation of a closed complex on substrate binding into a largely hydrophobic active site (8, 30, 31). Therefore, the inverse value for \( C_{14D} k_{non} \) may reflect conformational restriction of 4 within a low-volume active site of the Michaelis complex, resulting in a stiffer bonding environment for the C4–H center and an inverse binding isotope effect (32).

Conclusions
As the number of recognized \([4 + 2]\)-carbocycles continues to grow, experimental interrogation of reaction coordinates proposed from structural and computational studies will be necessary for a complete mechanistic understanding of these enzymes. The experimental results and computational predictions are collectively most consistent with the hypothesis that the nonenzymatic cyclization of 4 involves an intermediary state before complete C4–C12 bond formation. A possible candidate for such an intermediate (i.e., local free energy minimum) is the 10π-aromatic state (e.g., 7) identified in recent computational studies (12). The results also suggest that the SpnF-catalyzed reaction is similar to that of the nonenzymatic cyclization but likely involves a reduction in the entropy of activation after the substrate is bound in the active site, thereby facilitating catalysis and the observed rate enhancement (\( k_{enz}/k_{non} \)). These observations combined with prior structural investigations (8, 30, 31) suggest that \([4 + 2]\)-carbocycles in general may operate, in part, by reducing the conformation space available to the cyclizing species in a closed, induced fit complex, a criteria worthy.

\[ C_{TD} k_{non} \] and \( C_{11D} k_{non} \) are indicative of partial C7–C11 bond formation in the transition state of the most rate-limiting step. Likewise, the values centered near unity for \( C_{24D} k_{non} \) and \( C_{14D} k_{non} \) suggest little to no \( sp^2 \rightarrow sp^3 \) rehybridization at these positions in this transition state compared with the reference state (4). The ~4% normal value for \( C_{12D} k_{non} \) may be explained by redistribution of electron density along the C11–C15 π-system, leading to a loosening of the C12–H bonding environment (9, 11, 26).

These results are consistent with recent computational models of the reaction (12, 13, 27, 28). In particular, a simplified gas-phase model of the nonenzymatic reaction investigated by Hess and Smentek (13) identified a reaction coordinate with a single potential energy saddle point (similar to 6 in Fig. 2) followed by two apparent inflection points forming a caldera-like region. Subsequent computations by the Houk and Singleton laboratories using variational transition-state theory identified a local free energy minimum described as a 10π-aromatic species (depicted as structure 7 in Fig. 2) residing about the caldera between two variational transition states along the reaction coordinate (12). The first of these transition states (6) is similar to that identified by Hess and Smentek (13), and the trajectory of 7 can bifurcate on progression through the second transition state (8) to form either a \([4 + 2]\)- or \([6 + 4]\)-cycladduct (i.e., 5 vs. 9), where the latter then converts to 8 via a facile Cope rearrangement (12). The second of these transition states (8) is ~2.3 kcal/mol lower in free energy compared with the first (12). This implies that the sensitivity index at 300 K for the first transition state is nearly 50-fold greater than that for the second (24, 25). In other words, conversion of 4 to 7 via 6 is predicted to be significantly rate-limiting in the nonenzymatic reaction.

The KIEs calculated from Hess and Smentek’s gas-phase models of 6 vs. 4 (13) are also shown in Fig. 4 and are directionally consistent with the observed values. Because of uncertainty in the measurements, there are ~65 and 25% posterior probabilities that the \( C_{TD} k_{non} \) and \( C_{11D} k_{non} \) isotope effects, respectively, are between 0.90 and 0.95; however, the probability that they are less

Fig. 3. Modular, retrosynthetic design for the preparation of natural abundance and site-specifically deuterated isotopologs of 2. Carbons are numbered according to their final position in the macrolactone. Full synthetic schemata and protocols are provided in SI Appendix, SI Text and Figs. S5–S7.
of consideration for de novo protein design/reengineering to develop Diels–Alderases in the future.

Materials and Methods

Synthesis of Deuterated Compounds. Complete synthetic protocols and schemata (SI Appendix, Figs. S5–S7) for all deuterated compounds are provided in SI Appendix, SI Text.

Sample Preparation. The secondary deuterium KIE at each site of interest was measured using the method of internal competition (14) in at least three separate experimental trials. In each trial, site-specifically labeled and unlabeled SpnM substrates (3), which are stable, were mixed together (~1.5 mM total concentration) in 50 mM Tris buffer (pH 8.0) containing 10% DMSO at 30 °C before adding SpnM. Initial ratios of the isotopologs were generally close to 1 : 1, although they ranged from no less than 1 : 4 to no greater than 4 : 1. Reaction mixtures also included 0.47 mM p-methoxyacetophenone as an internal standard to standardize the observed HPLC peak integrations. The species 4 was generated in situ by the addition of SpnM to 16 μM, which converts all of the observable SpnM substrate (3) to SpnF substrate (4) within ~4 min. During this time, a byproduct of the SpnM reaction with the same mass as 4 also forms; however, controls showed that, after its formation, this by-product is both stable and baseline-separated from 4 by HPLC.

No SpnF was present during measurement of nonenzymatic KIEs, and reaction aliquots were taken after all of the SpnM substrate had been consumed at six different time points ranging from ~10 to 90% total reaction (~30 min time course). Proportionately larger aliquots were taken at later time points to compensate for the reduced concentration of residual substrate. During measurement of enzymatic KIEs, SpnF was added after all of the SpnM substrate had been consumed and immediately before taking the initial quench. SpnF was added in sufficient quantity, so that the final quench at ~80% total reaction was taken at ~6 min after adding SpnF (rationale is in SI Appendix, SI Text). In all other respects, the nonenzymatic and enzymatic samples were treated identically.

Isotopolog Enrichments. Isolated residual substrate was analyzed by MS at ~4% also forms; however, SI Appendix, SI Text 4 B in 3 min, isocratic at 70\% B, linear to 50\% B, linear to 50 \( \mu \)M SI: www.pnas.org/cgi/doi/10.1073/pnas.1710496114

Three 3-min continuous direct infusions were run in series for each sample. The first two 3-min infusions were used to flush the instrument to prevent memory effects. Data from these flushes were discarded. Signal intensities obtained from the third 3-min acquisition were averaged at each mass over 1-s intervals. Three 3-min continuous direct infusions were run in series for each sample. The first two 3-min infusions were used to flush the instrument to prevent memory effects. Data from these flushes were discarded. Signal intensities obtained from the third 3-min acquisition were averaged at each mass over 1-s intervals.

On removal, aliquots were quenched by mixing 1 : 1 with MeCN containing 5% DMSO at 0 °C, which arrests both enzymatic and nonenzymatic cyclization. Protein was removed by centrifugation, and the residual substrate (4) was isolated using C18 reversed-phase HPLC (Varian Microsorb-MV; 100-S C18 250 × 4.6 mm) with UV detection at 254 nm. The column was eluted with water (A) vs. MeCN (B) using the following gradient: start at 25% B, linear to 50% B in 35 min, linear to 70% B in 3 min, isocratic at 70% B for 2 min and linear return to 25% B in 2 min. A solvent blank was run interposed between each HPLC separation to prevent memory effects. Isolated 4 was concentrated in a speed vacuum concentrator and redissolved in a mixture of MeCN and water for MS analysis. Aliquots of the isolated substrate had been consumed and immediately before taking the initial quench. SpnF was added in sufficient quantity, so that the final quench at ~80% total reaction was taken at ~6 min after adding SpnF (rationale is in SI Appendix, SI Text). In all other respects, the nonenzymatic and enzymatic samples were treated identically.
**Statistical Analysis.** Each of the 12 measured KIEs was analyzed independently of the others using a hierarchical Bayesian regression model (33, 34). If n is the number of independent experimental trials for a given KIE, then the model involves 2n + 3 parameters. At the highest level is the KIE of interest denoted $\delta k$, for which a log-normal prior was constructed with median 0 and variance 0.0325, which reparameterizes to

$$
\rho (\delta k) = \frac{1}{\sqrt{2\pi}}N (\log \delta (0,0.0325))\delta k > 0.0.
$$

The distribution in Eq. 3 has a median of unity and ensures that the prior probability of $\delta k$ lying on a positive interval $[a, b]$ is equal to the prior probability that it is on $[1/b, 1/a]$. This can be interpreted physically as treating positive and negative differences in the free energy of activation between two isotopos equally. The variance of the log-normal distribution (0.0325) places ~68% of the prior probability mass between KIEs of ±0.6 and ±1.20, consistent with previous reports of $\delta$-secondary deuterium KIEs on rehybridization reactions at 300 K (9, 10, 29, 35–38). The prior also allows for the possibility of large KIEs, consistent with previous estimates of theoretical maxima (11, 29, 39) on the intervals [0.70, 0.83] and [1.20, 1.43] with a combined probability of 25%. Extreme KIEs that exceed the predicted maxima have a prior probability less than 5% with this prior. A sensitivity analysis described in SI Appendix, SI Text shows that the posterior inferences for $\delta k$ were insensitive to changes in the scale of this prior, unless it was increased so as to include extreme values with significant probability.

Between-experiment variation is modeled by characterizing the 7th trial with its own observed isotopic effect $\delta k_i$ and initial enrichment $\rho_i$. The prior probability distribution of $\rho_i$ is treated as uniform on (0, 1), consistent with the experimental design, and $\delta k$ is modeled as dependent on $\delta k_i$ according to

$$
\delta k_i = \delta k + \eta_i,
$$

where the error term $\eta_i$ is a truncated Gaussian random variable with scale $\sigma_\delta$ and mode 0. The prior distributions for $\delta k_i$ (conditional on $\delta k$ and $\sigma_\delta$) and $\rho_i$ are then given by

$$
\rho (\delta k_i | \delta k, \sigma_\delta) = N (\delta k, \sigma_\delta^2, \sigma_\delta^2), \delta k_i > 0,
$$

$$
\rho (\rho_i) = U (0, 1).
$$

Based on previous recommendations for hierarchical analyses with small numbers of groups (40), a half-Cauchy distribution on $(0, \infty)$ with scale 0.2 was used for $\rho (\sigma_\delta)$, which admits the possibility for large variation between experiments with significant probability. Sensitivity analysis indicated that doubling the scale from 0.2 to 0.4 had no meaningful impact on the posterior marginal probability intervals for $\delta k$. Finally, within-experiment variation is modeled as the SD about the regression, denoted $\sigma_w$, and the prior distribution for $\sigma_w$ is uniform on support (0, 1). The joint prior distribution is then given by

$$
\rho (\delta k, \sigma_b, \sigma_w, \delta k_i, \rho_i) = \rho (\delta k) \rho (\sigma_b) \rho (\sigma_w) \prod_{i=1}^n \rho (\delta k_i | \delta k, \sigma_\delta) \rho (\rho_i).
$$

The likelihood function is equal to the probability of the observed data conditional on the parameters. The likelihood function was thus constructed by modeling the 7th enrichment from the 7th experimental trial as

$$
\hat{f}_7 = R (f_7 \delta k_i, \rho_i) + \varepsilon_7,
$$

where the function $R$ is given by Eq. 1, $f_7$ is the 7th fraction of reaction in the 7th trial, and $\varepsilon_7$ is the 7th realization of the random Gaussian variable $\varepsilon$.


